HXT RDMA Enablement Based On HSRP v1.3/v1.4

Document Number: ???
Version: Rev 0.9
September 17,2018

Guizhou Huaxintong Semiconductor Technology Co., Ltd.
贵州华芯通半导体技术有限公司
Temporary Administrative Center, Guian New Area, Guizhou Province, P.R. China
贵州省贵安新区临时行政中心

Legal Notice

The document is:

- Confidential and proprietary to Guizhou Huaxintong Semiconductor Technology Co., Ltd. (“HXT Semiconductor”), and provided to designated receivers only, no public disclosure is permitted;
- Restricted to be distributed to anyone without the express approval of HXT Semiconductor;
- Not permitted to be used, copied, reproduced, modified, disclosed in whole or in part in any manner to others without the express written permission of HXT Semiconductor;

HUAXINTONG and its logo, are trademarks or registered trademarks of HXT Semiconductor in China. All other brand names, product names, or trademarks might belong to their respective holders.

HXT Semiconductor reserves the right to alter product and services offerings, and specifications and pricing at any time without notice, and is not responsible for typographical or graphical errors that may appear in this document.

The technical data included in this document may be subject to U.S. and international export, re-export, or transfer (“export”) laws. Diversion contrary to U.S. and international law is strictly prohibited.

1. 本文档中有关RDMA 的描述，参考了RDMA wiki 网页，其他所有内容属于原创。
2. 本文档中的所有内容，都可以按照华芯通知识产权随意传播，但是唯有测试出的benchmark 性能数据不能传播。因为没有经过系统调优，都是默认值，硬件和软件环境，各种连接方式，都会影响性能数据。

Revision History

Revision	Date	Description	Author
0.9	September. 21, 2018	Finished Soft RoCE, RoCE v1, RoCE v2, iWARP	Jianhua Xie

1. Introduction

1.1 Document Overview

This document is one of the document series on HXT Software Reference Platform (HSRP) v1.3 for the HXT Taishan Reference Design Platform (REP-1). REP-1 utilizes Qualcomm QDF2400 (HXT-1 product family compatible). This document lists the features of firmware/OS and applications, the released packages and the software development environment/target hardware platform. For additional information or technical questions, please contact HXT sales representative or support team.

1.2 Target Audience

This document is released in the hope of providing information to the software development engineers, platform designers and architectures and all other who is interested in HXT products under NDA.

This document doesn’t focus on performance benchmarking nor performance comparison.

2. RDMA Overview

RDMA - remote direct memory access is a direct memory access from the memory of one computer into that of another without involving either one's operating system. This permits high-throughput, low-latency networking, which is especially useful in massively parallel computer clusters. Please refer to this link:
https://en.wikipedia.org/wiki/Remote_direct_memory_access

2.1 RDMA implementations

There are some popular RDMA implementations, for example:

- RoCE - RDMA over Converged Ethernet - Mellanox solution, which also have some different version: v1, v2 and software based implementation – Soft RoCE.
- iWARP – Chelsio and Intel solution, recommended by IETF.
- InfiniBand – Mellanox solution

This documentation will only introduce Soft RoCE, RoCE v1, RoCE v2 and iWARP, since InfiniBand depends on Mellanox Dedicate Switch, and would be seldom deployed in new Data Centre cluster.

3. Test environment

3.1 Hardware – REP1

- Processor supported：Qualcomm QDF2400 Rev2.0 (HXT-1 product family compatible)
  - CPU operating frequency 2500 MHz
  - CBF operating frequency 2100 MHz
  - DDR4 – default, no change
- Platform supported: HXT Taishan Reference Design Platform (REP-1)
- NICs
  - Mellanox ConnectX-413A 56/40GbE, PCIE Gen3 x8 card in x16 slot
  - Chelsio Communications Inc T6225-CR 25GbE, PCIE Gen3 x8
  - Intel Corporation 82599ES 10-GbE, PCIE Gen3 x8
- NIC connectivity

  All NIC connections are back-to-back by direct cables, no Switch involved.

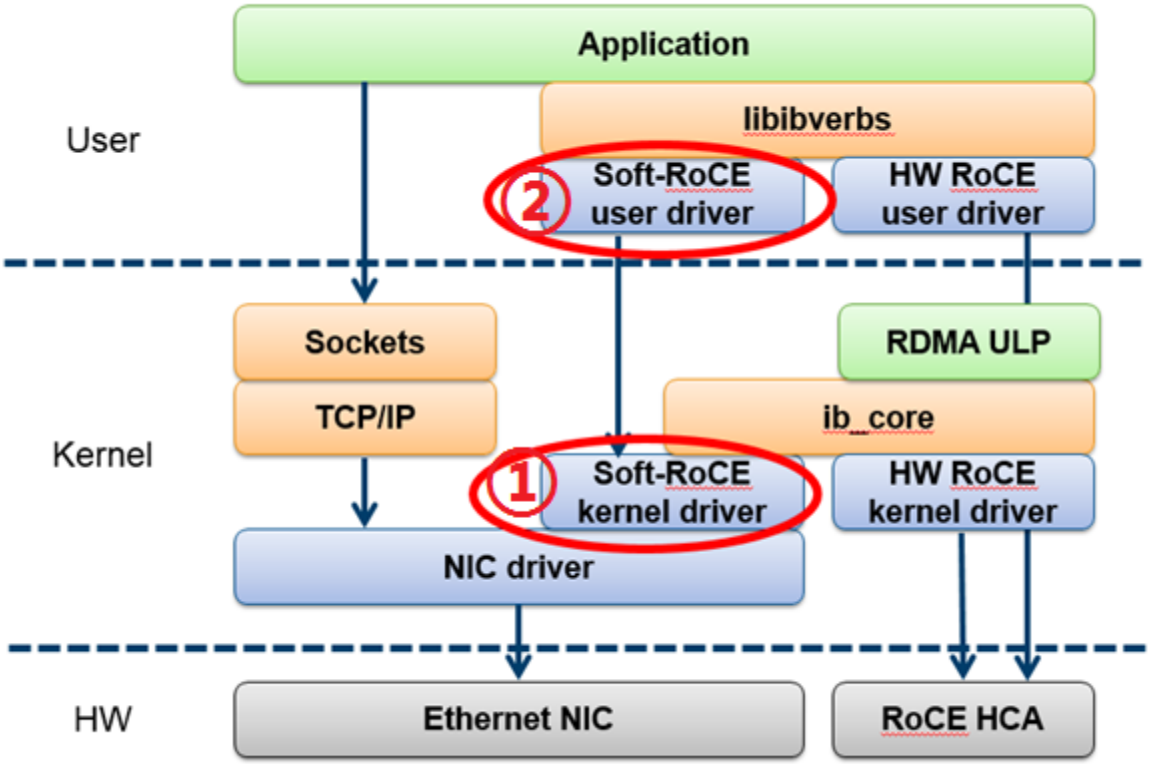
3.2 Software – HSRP v1.3/v1.4

- Linux Kernel Version: kernel 4.14.15.hxt.aarch64, 4.14. 36-6.hxt.aarch64
- Linux Distro Version: CentOS 7.4
- BMC – default, no change
- UEFI – default, no change
- NIC Driver:
  - Mellanox OFED4.3
  - Intel ixgbe version: 5.1.0-k
  - ChelsioUwire-3.8.0.2

4. Test procedure

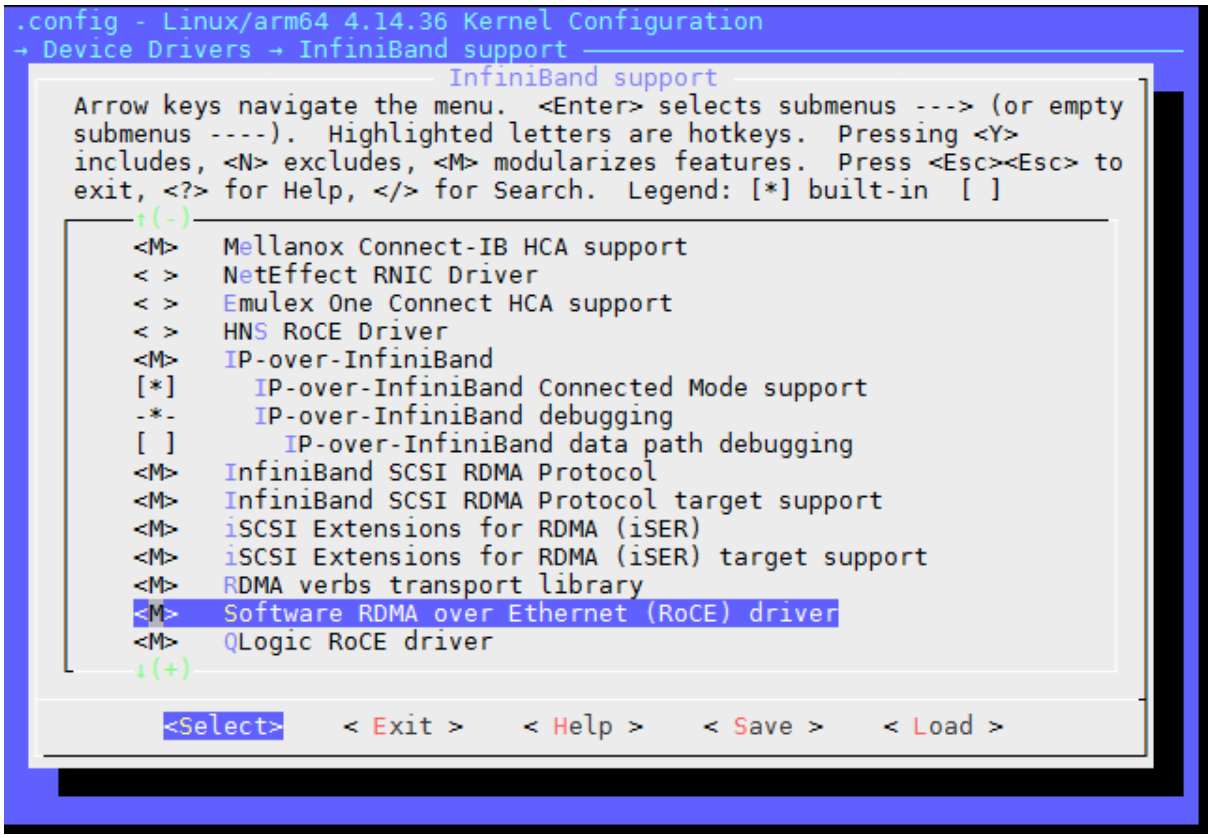
4.1 Soft RoCE

The architecture of Soft-RoCE is as below:



4.1.1 Install kernel driver

At the website of Mellanox, there is a Mellanox documentation "HowTo Configure Soft-RoCE", which was finished in 2015. However, part 1 in above picture, the kernel driver of Soft-RoCE was merged into the Linux mainline in Jun 2016 by Moni Shoua monis@mellanox.com. So this git repository https://github.com/SoftRoCE/rxe-dev.git is not needed any more. To enable this kernel driver on HXT ARM64 platform, the rxe option should be selected and be built as below:



4.1.2 Install user space libraries

Part 2 in the architecture of Soft-RoCE, which was tracking by a git repository: https://github.com/SoftRoCE/librxe-dev.git. However, in HSRP v1.4 software environment, while those common RDMA tools are installed, this user space library will also be installed automatically as dependencies.

4.1.3 Install user space tools

Install below user space tools with yum command:

#yum install -y automake bc elfutils-libelf-devel epel-release gcc gcc-c++ libibverbs libibverbs-devel libibverbs-utils libnl-devel libnl3-devel librdmacm librdmacm-devel librdmacm-utils ncurses-devel openssl-devel perftest perl-Switch valgrind-devel

4.1.4 Test steps

1. Load kernel driver ib_core ib_uverbs rdma_ucm rdma_rxe by performing command at both server peer and client peer:
#rxe_cfg start

2. Check NICs link status at the server peer:
rxe_cfg status

```
[root@chelsio-102 ~]# rxe_cfg status
  Name          Link  Driver      Speed   NMTU  IPv4_addr  RDEV  RMTU
  dummy0              dummy
  enp1s0        yes   mlx5_core
  enP4p1s0f0    yes   ixgbe       10GigE
  enP4p1s0f1    no    ixgbe       10GigE
  enP5p1s0f4    yes   cxgb4       10GigE
  enP5p1s0f4d1  no    cxgb4       10GigE
  eth0          yes   qcom-emac
[root@chelsio-102 ~]#
```

Mellanox CX-413A
Intel X540-T2
Chelsio T6225-CR

3. Check NICs link status at the client peer:

rxe_cfg status

```
[root@chelsio-104 ~]# rxe_cfg status
  Name          Link  Driver      Speed   NMTU  IPv4_addr  RDEV  RMTU
  dummy0              dummy
  enp1s0        yes   mlx5_core
  enP4p1s0f4    yes   cxgb4       10GigE
  enP4p1s0f4d1  no    cxgb4       10GigE
  enP5p1s0f0    yes   ixgbe       10GigE
  enP5p1s0f1    no    ixgbe       10GigE
  eth0          yes   qcom-emac
[root@chelsio-104 ~]#
```

Mellanox CX-413A
Intel X540-T2
Chelsio T6225-CR

4. Add new rxe device to 3 Ethernet NIC at the server peer, then double-check them:

rxe_cfg add enp1s0
rxe_cfg add enP4p1s0f0
rxe_cfg add enP5p1s0f4
rxe_cfg status

```
[root@chelsio-102 ~]# rxe_cfg status
  Name          Link  Driver      Speed   NMTU  IPv4_addr  RDEV  RMTU
  dummy0              dummy
  enp1s0        yes   mlx5_core                            rxe0  1024  (3)
  enP4p1s0f0    yes   ixgbe       10GigE                   rxe1  1024  (3)
  enP4p1s0f1    no    ixgbe       10GigE
  enP5p1s0f4    yes   cxgb4       10GigE                   rxe2  1024  (3)
  enP5p1s0f4d1  no    cxgb4       10GigE
  eth0          yes   qcom-emac
[root@chelsio-102 ~]#
```

5. Add new rxe device to 3 Ethernet NIC at the client peer , then double-check them:

rxe_cfg add enp1s0
rxe_cfg add enP5p1s0f0
rxe_cfg add enP4p1s0f4
rxe_cfg status

```
[root@chelsio-104 ~]# rxe_cfg status
  Name          Link  Driver     Speed   NMTU  IPv4_addr  RDEV  RMTU
  dummy0              dummy
  enp1s0        yes   mlx5_core                           rxe0  1024  (3)
  enP4p1s0f4    yes   cxgb4      10GigE                   rxe2  1024  (3)
  enP4p1s0f4d1  no    cxgb4      10GigE
  enP5p1s0f0    yes   ixgbe      10GigE                   rxe1  1024  (3)
  enP5p1s0f1    no    ixgbe      10GigE
  eth0          yes   qcom-emac
[root@chelsio-104 ~]#
```

6. Configure NICs with IPv4 address at the server peer:

#ifconfig enp1s0 192.85.0.1/24 up
#ifconfig enP5p1s0f4 192.86.1.1/24 up
#ifconfig enP4p1s0f0 10.10.10.1/24 up

7. Configure NICs with IPv4 address at the client peer:

ifconfig enp1s0 192.85.0.2/24 up
ifconfig enP4p1s0f4 192.86.1.2/24 up
ifconfig enP5p1s0f0 10.10.10.2/24 up

8. Run rping at server peer to test Mellanox NIC:

rping -s -a 192.85.0.1 -v -C 10

9. Run rping at client peer to test Mellanox NIC:

#rping -c -a 192.85.0.1 -v -C 10

10. Server peer shows the Mellanox NIC rping result as below:

```
[root@chelsio-102 ~]# rping -s -a 192.85.0.1 -v -C 10
server ping data: rdma-ping-0: ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqr
server ping data: rdma-ping-1: BCDEFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrs
server ping data: rdma-ping-2: CDEFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrst
server ping data: rdma-ping-3: DEFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstu
server ping data: rdma-ping-4: EFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuv
server ping data: rdma-ping-5: FGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvw
server ping data: rdma-ping-6: GHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwx
server ping data: rdma-ping-7: HIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxy
server ping data: rdma-ping-8: IJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyz
server ping data: rdma-ping-9: JKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyzA
server DISCONNECT EVENT...
wait for RDMA_READ_ADV state 10
[root@chelsio-102 ~]#
```

11. Client peer shows the Mellanox NIC rping result as below:

```
[root@chelsio-104 ~]# rping -c -a 192.85.0.1 -v -C 10
ping data: rdma-ping-0: ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqr
ping data: rdma-ping-1: BCDEFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrs
ping data: rdma-ping-2: CDEFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrst
ping data: rdma-ping-3: DEFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstu
ping data: rdma-ping-4: EFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuv
ping data: rdma-ping-5: FGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvw
ping data: rdma-ping-6: GHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwx
ping data: rdma-ping-7: HIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxy
ping data: rdma-ping-8: IJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyz
ping data: rdma-ping-9: JKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyzA
client DISCONNECT EVENT...
[root@chelsio-104 ~]#
```

12. Run rping at server peer to test Chelsio NIC:
rping -s -a 192.86.1.1 -v -C 10

13. Run rping at client peer to test Chelsio NIC:
rping -c -a 192.86.1.1 -v -C 10

14. Server peer shows the Chelsio NIC rping result as below:

```
[root@chelsio-102 ~]# ifconfig enP5p1s0f4 192.86.1.1/24 up
[root@chelsio-102 ~]# rping -s -a 192.86.1.1 -v -C 10
server ping data: rdma-ping-0: ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqr
server ping data: rdma-ping-1: BCDEFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrs
server ping data: rdma-ping-2: CDEFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrst
server ping data: rdma-ping-3: DEFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstu
server ping data: rdma-ping-4: EFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuv
server ping data: rdma-ping-5: FGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvw
server ping data: rdma-ping-6: GHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwx
server ping data: rdma-ping-7: HIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxy
server ping data: rdma-ping-8: IJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyz
server ping data: rdma-ping-9: JKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyzA
server DISCONNECT EVENT...
wait for RDMA_READ_ADV state 10
[root@chelsio-102 ~]#
```

15. Client peer shows the Chelsio NIC rping result as below:

```
[root@chelsio-104 ~]# ifconfig enP4p1s0f4 192.86.1.2/24 up
[root@chelsio-104 ~]# rping -c -a 192.86.1.1 -v -C 10
ping data: rdma-ping-0: ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqr
ping data: rdma-ping-1: BCDEFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrs
ping data: rdma-ping-2: CDEFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrst
ping data: rdma-ping-3: DEFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstu
ping data: rdma-ping-4: EFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuv
ping data: rdma-ping-5: FGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvw
ping data: rdma-ping-6: GHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwx
ping data: rdma-ping-7: HIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxy
ping data: rdma-ping-8: IJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyz
ping data: rdma-ping-9: JKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyzA
client DISCONNECT EVENT...
[root@chelsio-104 ~]#
```

16. Run rping at server peer to test Intel NIC:

rping -s -a 10.10.10.1 -v -C 10

17. Run rping at client peer to test Intel NIC:
rping -c -a 10.10.10.1 -v -C 10

18. Server peer shows the Intel NIC rping result as below:

```
[root@chelsio-102 ~]# ifconfig enP4p1s0f0 10.10.10.1/24 up
[root@chelsio-102 ~]# rping -s -a 10.10.10.1 -v -C 10
server ping data: rdma-ping-0: ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqr
server ping data: rdma-ping-1: BCDEFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrs
server ping data: rdma-ping-2: CDEFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrst
server ping data: rdma-ping-3: DEFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstu
server ping data: rdma-ping-4: EFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuv
server ping data: rdma-ping-5: FGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvw
server ping data: rdma-ping-6: GHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwx
server ping data: rdma-ping-7: HIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxy
server ping data: rdma-ping-8: IJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyz
server ping data: rdma-ping-9: JKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyzA
server DISCONNECT EVENT...
wait for RDMA_READ_ADV state 10
[root@chelsio-102 ~]#
```

19. Client peer shows the Intel NIC rping result as below:

```
[root@chelsio-104 ~]# ifconfig enP5p1s0f0 10.10.10.2/24 up
[root@chelsio-104 ~]# rping -c -a 10.10.10.1 -v -C 10
ping data: rdma-ping-0: ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqr
ping data: rdma-ping-1: BCDEFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrs
ping data: rdma-ping-2: CDEFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrst
ping data: rdma-ping-3: DEFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstu
ping data: rdma-ping-4: EFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuv
ping data: rdma-ping-5: FGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvw
ping data: rdma-ping-6: GHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwx
ping data: rdma-ping-7: HIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxy
ping data: rdma-ping-8: IJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyz
ping data: rdma-ping-9: JKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyzA
client DISCONNECT EVENT...
[root@chelsio-104 ~]#
```

20. Test write latency with Mellanox NIC at server peer:
ib_write_lat -d rxe0 -i 1 -a -R

21. Test write latency with Mellanox NIC at client peer:
ib_write_lat 192.85.0.1 -a -d rxe0 -i 1 –R

22. Server peer shows the Mellanox NIC write latency result as below:

```
[root@chelsio-102 ~]# ib_write_lat -d rxe0 -i 1 -a -R

************************************
* Waiting for client to connect... *
************************************
---------------------------------------------------------------------------------------
                    RDMA_Write Latency Test
 Dual-port       : OFF          Device         : rxe0
 Number of qps   : 1            Transport type : IB
 Connection type : RC           Using SRQ      : OFF
 Mtu             : 1024[B]
 Link type       : Ethernet
 GID index       : 2
 Max inline data : 0[B]
 rdma_cm QPs     : ON
 Data ex. method : rdma_cm
---------------------------------------------------------------------------------------
 Waiting for client rdma_cm QP to connect
 Please run the same command with the IB/RoCE interface IP
---------------------------------------------------------------------------------------
 local address: LID 0000 QPN 0xc0d3 PSN 0xac9e9e
 GID: 00:00:00:00:00:00:00:00:00:00:255:255:192:85:00:01
 remote address: LID 0000 QPN 0x00d3 PSN 0x4b09dc
 GID: 00:00:00:00:00:00:00:00:00:00:255:255:192:85:00:02
---------------------------------------------------------------------------------------
 #bytes #iterations    t_min[usec]    t_max[usec]  t_typical[usec]    t_avg[usec]    t_stdev[usec]   99% percentile[usec]   99.9% percentile[usec]
 2       1000          2.12           2.25         2.15               2.15           0.02            2.22                   2.25
 4       1000          2.12           6.30         2.15               2.15           0.02            2.20                   6.30
 8       1000          2.12           5.30         2.15               2.15           0.03            2.20                   5.30
 16      1000          2.12           2.25         2.15               2.15           0.02            2.20                   2.25
 32      1000          2.12           2.22         2.15               2.16           0.01            2.20                   2.22
 64      1000          2.15           6.40         2.17               2.19           0.04            2.22                   6.40
 128     1000          2.20           3.00         2.25               2.24           0.02            2.27                   3.00
 256     1000          2.30           2.40         2.32               2.33           0.01            2.37                   2.40
 512     1000          2.42           3.42         2.47               2.48           0.02            2.52                   3.42
 1024    1000          2.65           4.15         2.67               2.69           0.06            2.72                   4.15
 2048    1000          2.87           2.97         2.92               2.91           0.02            2.95                   2.97
 4096    1000          3.20           6.85         3.25               3.24           0.03            3.30                   6.85
 8192    1000          3.85           4.72         3.90               3.90           0.02            3.95                   4.72
 16384   1000          5.22           7.22         5.27               5.30           0.09            5.57                   7.22
 32768   1000          8.82           9.22         9.05               9.06           0.07            9.20                   9.22
 65536   1000          16.45          17.17        16.57              16.68          0.18            17.10                  17.17
 131072  1000          26.82          28.02        26.92              26.94          0.08            27.22                  28.02
 262144  1000          47.60          50.55        47.67              47.72          0.12            48.05                  50.55
 524288  1000          89.12          90.09        89.22              89.28          0.13            89.64                  90.09
 1048576 1000          172.18         177.11       172.28             172.41         0.23            172.93                 177.11
 2097152 1000          338.29         339.32       338.69             338.70         0.26            339.22                 339.32
 4194304 1000          670.49         671.64       670.97             670.97         0.30            671.52                 671.64
 8388608 1000          1334.91        1336.16      1335.31            1335.29        0.29            1335.93                1336.16
---------------------------------------------------------------------------------------
[root@chelsio-102 ~]#
```

23. client peer shows the Mellanox NIC write latency result as below:

```
[root@chelsio-104 ~]# ib_write_lat 192.85.0.1 -a -d rxe0 -i 1 -R
---------------------------------------------------------------------------------------
                    RDMA_Write Latency Test
 Dual-port       : OFF          Device         : rxe0
 Number of qps   : 1            Transport type : IB
 Connection type : RC           Using SRQ      : OFF
 TX depth        : 1
 Mtu             : 1024[B]
 Link type       : Ethernet
 GID index       : 2
 Max inline data : 0[B]
 rdma_cm QPs     : ON
 Data ex. method : rdma_cm
---------------------------------------------------------------------------------------
 local address: LID 0000 QPN 0x00d3 PSN 0x4b09dc
 GID: 00:00:00:00:00:00:00:00:00:00:255:255:192:85:00:02
 remote address: LID 0000 QPN 0x00d3 PSN 0xac9e9e
 GID: 00:00:00:00:00:00:00:00:00:00:255:255:192:85:00:01
---------------------------------------------------------------------------------------
 #bytes #iterations    t_min[usec]    t_max[usec]  t_typical[usec]    t_avg[usec]    t_stdev[usec]   99% percentile[usec]   99.9% percentile[usec]
 2       1000          2.12           2.22         2.15               2.15           0.02            2.20                   2.22
 4       1000          2.12           2.87         2.15               2.15           0.01            2.20                   2.87
 8       1000          2.12           5.12         2.15               2.15           0.02            2.20                   5.12
 16      1000          2.12           2.22         2.15               2.15           0.01            2.20                   2.22
 32      1000          2.12           2.20         2.15               2.16           0.01            2.20                   2.20
 64      1000          2.15           6.35         2.17               2.19           0.02            2.22                   6.35
 128     1000          2.20           2.60         2.25               2.24           0.02            2.27                   2.60
 256     1000          2.30           2.37         2.32               2.33           0.01            2.37                   2.37
 512     1000          2.45           2.92         2.47               2.48           0.01            2.50                   2.92
 1024    1000          2.65           4.15         2.67               2.69           0.05            2.72                   4.15
 2048    1000          2.87           3.00         2.92               2.91           0.02            2.95                   3.00
 4096    1000          3.20           6.95         3.25               3.24           0.02            3.27                   6.95
 8192    1000          3.87           4.35         3.90               3.90           0.02            3.95                   4.35
 16384   1000          5.22           6.17         5.27               5.30           0.10            5.57                   6.17
 32768   1000          8.82           9.22         9.05               9.06           0.07            9.20                   9.22
 65536   1000          16.45          17.05        16.77              16.68          0.14            16.92                  17.05
 131072  1000          26.85          27.65        26.92              26.94          0.08            27.25                  27.65
 262144  1000          47.60          50.57        47.67              47.72          0.12            48.10                  50.57
 524288  1000          89.12          89.74        89.22              89.28          0.14            89.64                  89.74
 1048576 1000          172.19         177.11       172.36             172.42         0.20            172.76                 177.11
 2097152 1000          338.29         339.27       338.69             338.70         0.19            339.09                 339.27
 4194304 1000          670.51         671.61       670.94             670.96         0.31            671.54                 671.61
 8388608 1000          1334.89        1336.04      1335.16            1335.29        0.32            1335.96                1336.04
---------------------------------------------------------------------------------------
[root@chelsio-104 ~]#
```

24. Test write latency with Chelsio NIC at server peer:
ib_write_lat -d rxe1 -i 1 -a -R

25. Test write latency with Chelsio NIC at client peer:
ib_write_lat 192.86.1.1 -a -d rxe1 -i 1 –R

26. Server peer shows the Chelsio NIC write latency result as below:

```
[root@chelsio-102 ~]# ib_write_lat -d rxe1 -i 1 -a -R

************************************
* Waiting for client to connect... *
************************************
---------------------------------------------------------------------------------------
                    RDMA_Write Latency Test
 Dual-port       : OFF          Device         : rxe1
 Number of qps   : 1            Transport type : IB
 Connection type : RC           Using SRQ      : OFF
 Mtu             : 1024[B]
 Link type       : Ethernet
 GID index       : 0
 Max inline data : 0[B]
 rdma_cm QPs     : ON
 Data ex. method : rdma_cm
---------------------------------------------------------------------------------------
 Waiting for client rdma_cm QP to connect
 Please run the same command with the IB/RoCE interface IP
---------------------------------------------------------------------------------------
 local address: LID 0000 QPN 0x0502 PSN 0x11d9e
 GID: 00:07:67:62:138:240:00:00:00:00:00:00:00:00:00:00
 remote address: LID 0000 QPN 0x0502 PSN 0xe28176
 GID: 00:07:67:62:139:80:00:00:00:00:00:00:00:00:00:00
---------------------------------------------------------------------------------------
 #bytes #iterations    t_min[usec]    t_max[usec]  t_typical[usec]    t_avg[usec]    t_stdev[usec]   99% percentile[usec]   99.9% percentile[usec]
 2       1000          4.20           5.72         4.22               4.24           0.04            4.40                   5.72
 4       1000          4.17           5.60         4.22               4.23           0.05            4.40                   5.60
 8       1000          4.17           6.35         4.22               4.24           0.07            4.35                   6.35
 16      1000          4.20           5.95         4.25               4.25           0.05            4.37                   5.95
 32      1000          4.20           5.02         4.25               4.27           0.04            4.40                   5.02
 64      1000          4.25           8.45         4.30               4.31           0.06            4.42                   8.45
 128     1000          4.35           8.07         4.40               4.41           0.07            4.55                   8.07
 256     1000          4.52           6.92         4.75               4.76           0.12            5.05                   6.92
 512     1000          4.85           7.35         5.25               5.25           0.09            5.40                   7.35
 1024    1000          5.32           9.65         5.55               5.55           0.10            5.70                   9.65
 2048    1000          5.75           8.05         5.92               5.92           0.05            6.05                   8.05
 4096    1000          6.37           10.20        6.45               6.45           0.06            6.55                   10.20
 8192    1000          7.75           8.70         7.82               7.81           0.02            7.85                   8.70
 16384   1000          10.62          12.30        10.65              10.67          0.06            10.82                  12.30
 32768   1000          16.30          18.05        16.35              16.35          0.05            16.47                  18.05
 65536   1000          27.52          30.95        27.57              27.58          0.10            27.70                  30.95
 131072  1000          49.95          50.87        50.00              50.01          0.04            50.15                  50.87
 262144  1000          94.79          95.89        94.84              94.86          0.05            95.07                  95.89
 524288  1000          184.58         185.48       184.68             184.68         0.05            184.83                 185.48
 1048576 1000          364.07         365.04       364.14             364.14         0.05            364.27                 365.04
 2097152 1000          723.13         724.33       723.21             723.22         0.05            723.33                 724.33
 4194304 1000          1441.29        1441.54      1441.37            1441.38        0.05            1441.52                1441.54
 8388608 1000          2877.48        2878.51      2877.56            2877.58        0.05            2877.71                2878.51
---------------------------------------------------------------------------------------
[root@chelsio-102 ~]#
```

27. client peer shows the Chelsio NIC write latency result as below:

```
[root@chelsio-104 ~]# ib_write_lat 192.86.1.1 -a -d rxe1 -i 1 -R
---------------------------------------------------------------------------------------
                    RDMA_Write Latency Test
 Dual-port       : OFF          Device         : rxe1
 Number of qps   : 1            Transport type : IB
 Connection type : RC           Using SRQ      : OFF
 TX depth        : 1
 Mtu             : 1024[B]
 Link type       : Ethernet
 GID index       : 0
 Max inline data : 0[B]
 rdma_cm QPs     : ON
 Data ex. method : rdma_cm
---------------------------------------------------------------------------------------
 local address: LID 0000 QPN 0x0502 PSN 0xe28176
 GID: 00:07:67:62:139:80:00:00:00:00:00:00:00:00:00:00
 remote address: LID 0000 QPN 0x0502 PSN 0x11d9e
 GID: 00:07:67:62:138:240:00:00:00:00:00:00:00:00:00:00
---------------------------------------------------------------------------------------
 #bytes #iterations    t_min[usec]    t_max[usec]  t_typical[usec]    t_avg[usec]    t_stdev[usec]   99% percentile[usec]   99.9% percentile[usec]
 2       1000          4.20           5.10         4.22               4.24           0.04            4.37                   5.10
 4       1000          4.17           5.27         4.22               4.23           0.04            4.35                   5.27
 8       1000          4.17           6.52         4.22               4.24           0.06            4.35                   6.52
 16      1000          4.20           5.72         4.25               4.25           0.04            4.37                   5.72
 32      1000          4.22           4.70         4.25               4.27           0.04            4.40                   4.70
 64      1000          4.25           8.27         4.30               4.31           0.05            4.42                   8.27
 128     1000          4.35           8.27         4.40               4.41           0.07            4.55                   8.27
 256     1000          4.52           6.90         4.75               4.76           0.12            5.02                   6.90
 512     1000          4.77           6.35         5.25               5.25           0.08            5.40                   6.35
 1024    1000          5.30           9.65         5.55               5.55           0.10            5.72                   9.65
 2048    1000          5.75           8.12         5.90               5.92           0.05            6.05                   8.12
 4096    1000          6.40           10.15        6.45               6.45           0.05            6.55                   10.15
 8192    1000          7.75           8.22         7.82               7.81           0.02            7.85                   8.22
 16384   1000          10.62          12.27        10.65              10.67          0.06            10.82                  12.27
 32768   1000          16.30          18.00        16.35              16.35          0.04            16.47                  18.00
 65536   1000          27.52          30.47        27.57              27.58          0.08            27.70                  30.47
 131072  1000          49.95          50.45        50.00              50.01          0.05            50.15                  50.45
 262144  1000          94.79          95.94        94.84              94.85          0.05            95.04                  95.94
 524288  1000          184.61         185.48       184.68             184.68         0.04            184.83                 185.48
 1048576 1000          364.07         364.99       364.14             364.14         0.04            364.24                 364.99
 2097152 1000          723.14         724.36       723.21             723.21         0.04            723.34                 724.36
 4194304 1000          1441.29        1441.54      1441.37            1441.37        0.04            1441.49                1441.54
 8388608 1000          2877.51        2878.08      2877.58            2877.59        0.05            2877.73                2878.08
---------------------------------------------------------------------------------------
[root@chelsio-104 ~]# ifconfig
```

28. Test write latency with Intel NIC at server peer:
ib_write_lat -d rxe2 -i 1 -a -R

29. Test write latency with Intel NIC at client peer:
ib_write_lat 10.10.10.1 -a -d rxe2 -i 1 -R

30. Server peer shows the Intel NIC write latency result as below:

```
[root@chelsio-102 ~]# ib_write_lat -d rxe2 -i 1 -a -R

************************************
* Waiting for client to connect... *
************************************
---------------------------------------------------------------------------------------
                    RDMA_Write Latency Test
 Dual-port       : OFF          Device         : rxe2
 Number of qps   : 1            Transport type : IB
 Connection type : RC           Using SRQ      : OFF
 Mtu             : 1024[B]
 Link type       : Ethernet
 GID index       : 1
 Max inline data : 0[B]
 rdma_cm QPs     : ON
 Data ex. method : rdma_cm
---------------------------------------------------------------------------------------
 Waiting for client rdma_cm QP to connect
 Please run the same command with the IB/RoCE interface IP
---------------------------------------------------------------------------------------
 local address: LID 0000 QPN 0x0012 PSN 0x70c6b6
 GID: 00:00:00:00:00:00:00:00:00:00:255:255:10:10:10:01
 remote address: LID 0000 QPN 0x0012 PSN 0xa01028
 GID: 00:00:00:00:00:00:00:00:00:00:255:255:10:10:10:02
---------------------------------------------------------------------------------------
 #bytes #iterations    t_min[usec]    t_max[usec]  t_typical[usec]    t_avg[usec]    t_stdev[usec]   99% percentile[usec]   99.9% percentile[usec]
 2       1000          23.85          76.49        25.10              25.29          2.81            26.37                  76.49
 4       1000          19.17          31.22        25.02              24.94          0.50            25.87                  31.22
 8       1000          23.97          27.00        25.05              25.00          0.33            26.42                  27.00
 16      1000          23.00          28.80        25.10              25.08          0.29            26.27                  28.80
 32      1000          19.45          76.89        25.05              25.06          1.20            26.85                  76.89
 64      1000          18.52          28.22        25.07              24.98          0.88            27.72                  28.22
 128     1000          22.00          27.12        25.47              25.39          0.40            26.57                  27.12
 256     1000          19.55          53.19        25.50              25.46          1.35            26.42                  53.19
 512     1000          20.90          59.79        23.57              23.93          1.94            26.00                  59.79
 1024    1000          25.67          35.72        28.62              28.33          0.54            29.42                  35.72
 2048    1000          25.60          78.84        40.95              41.00          2.10            43.80                  78.84
 4096    1000          37.92          82.02        40.97              41.65          5.00            80.69                  82.02
 8192    1000          60.02          296.55       115.99             130.07         52.32           276.60                 296.55
 16384   1000          100.64         376.17       178.88             173.30         55.35           312.47                 376.17
 32768   1000          182.86         450.86       299.80             292.65         67.09           429.19                 450.86
 65536   1000          341.09         1040.86      514.88             475.94         94.60           756.78                 1040.86
 131072  1000          682.09         855.27       695.81             699.43         13.94           731.91                 855.27
 262144  1000          1014.76        2132.26      1341.58            1294.66        113.10          1386.12                2132.26
 524288  1000          1288.35        1099898.70       1443.51        6948.68        76637.72         2306.36                1099898.70
 1048576 1000          2379.19        1101536.53       2454.84        7011.90        68778.71         4682.52                1101536.53
 2097152 1000          4514.58        1094486.31       5120.19        14335.27       96340.87         9729.14                1094486.31
 4194304 1000          9065.62        1116775.98       13309.67               17046.46        68606.02        18480.19               1116775.98
 8388608 1000          18136.22        1130887.83       27428.01              28936.01        49302.57        37796.82               1130887.83
---------------------------------------------------------------------------------------
[root@chelsio-102 ~]#
```

31. client peer shows the Intel NIC write latency result as below:

```
[root@chelsio-104 ~]# ib_write_lat 10.10.10.1 -a -d rxe2 -i 1 -R
---------------------------------------------------------------------------------------
                    RDMA_Write Latency Test
 Dual-port       : OFF          Device         : rxe2
 Number of qps   : 1            Transport type : IB
 Connection type : RC           Using SRQ      : OFF
 TX depth        : 1
 Mtu             : 1024[B]
 Link type       : Ethernet
 GID index       : 1
 Max inline data : 0[B]
 rdma_cm QPs     : ON
 Data ex. method : rdma_cm
---------------------------------------------------------------------------------------
 local address: LID 0000 QPN 0x0012 PSN 0xa01028
 GID: 00:00:00:00:00:00:00:00:00:00:255:255:10:10:10:02
 remote address: LID 0000 QPN 0x0012 PSN 0x70c6b6
 GID: 00:00:00:00:00:00:00:00:00:00:255:255:10:10:10:01
---------------------------------------------------------------------------------------
 #bytes #iterations    t_min[usec]    t_max[usec]  t_typical[usec]    t_avg[usec]    t_stdev[usec]   99% percentile[usec]   99.9% percentile[usec]
 2       1000          21.87          76.09        25.10              25.26          2.41            26.20                  76.09
 4       1000          20.47          28.85        25.00              24.95          0.43            26.07                  28.85
 8       1000          24.12          26.32        25.05              25.00          0.33            25.95                  26.32
 16      1000          20.30          28.75        25.07              25.08          0.33            26.72                  28.75
 32      1000          20.27          76.02        25.02              25.06          1.14            26.57                  76.02
 64      1000          19.60          27.77        25.07              24.98          0.85            27.17                  27.77
 128     1000          19.80          27.95        25.50              25.39          0.55            26.62                  27.95
 256     1000          19.72          56.12        25.50              25.46          1.38            26.55                  56.12
 512     1000          20.65          55.12        23.57              23.92          1.84            25.77                  55.12
 1024    1000          24.67          39.42        28.62              28.33          0.57            29.32                  39.42
 2048    1000          29.82          79.12        40.95              41.01          1.92            43.22                  79.12
 4096    1000          38.10          84.09        40.97              41.64          5.05            80.59                  84.09
 8192    1000          59.49          313.12       103.49             130.06         60.75           280.67                 313.12
 16384   1000          98.37          331.57       135.06             173.35         66.07           327.52                 331.57
 32768   1000          182.03         469.28       237.30             292.63         89.26           453.86                 469.28
 65536   1000          346.37         908.22       396.79             476.36         124.58          754.81                 908.22
 131072  1000          680.72         852.25       695.44             699.41         13.93           732.49                 852.25
 262144  1000          1023.71        2140.56      1341.78            1294.49        113.42          1390.18                2140.56
 524288  1000          1294.35        1100537.45       1441.52        6947.71        76646.13         2244.94                1100537.45
 1048576 1000          2371.85        1101584.60       2458.17        7012.17        68793.24         4666.44                1101584.60
 2097152 1000          4552.71        1097156.94       5101.56        14329.18       96358.09         9186.99                1097156.94
 4194304 1000          9076.04        1120331.80       13300.05               17042.49        68789.07        17841.36               1120331.80
 8388608 1000          18110.48        1132541.36       27299.71              28926.15        49423.44        37494.83               1132541.36
---------------------------------------------------------------------------------------
[root@chelsio-104 ~]#
```

32. Question: it seems that the Intel NICs has the worst write latency?

Answer: Mellanox CX-413A has 50Gbps link speed, Chelsio T6225-CR has 25Gbps

link speed, but Intel X540-T2 has only 10Gbps link speed. Besides, 2 Mellanox CX-413A connect by a direct cable, 2 Chelsio T6225-CR connect by a direct cable, but Intel X540-T2 NICs connect a L1 Switch.

4.2 RoCE v1/v2

4.2.1 Install kernel header files

#yum install kernel-devel-4.14.36-6.hxt.aarch64.rpm

4.2.2 Install kernel driver dependencies

#yum install -y python-devel redhat-rpm-config rpm-build gcc createrepo gtk2 atk cairo tcl gcc-gfortran tk

4.2.3 Install NIC kernel driver and user space tools

#./mlnxofedinstall --add-kernel-support --skip-distro-check

4.2.4 Test steps

- Configure NIC with an IPv4 address

#ifconfig enp1s0 192.85.1.2/24 up

- Show gids and RoCE version at the server peer

```
[root@hsrp1 ~]# ifconfig enp1s0 192.85.1.1/24 up
[root@hsrp1 ~]# show_gids
DEV     PORT    INDEX   GID                                     IPv4            VER     DEV
---     ----    -----   ---                                     ------------    ---     ---
mlx5_0  1       0       fe80:0000:0000:0000:268a:07ff:feb5:6ef4                 v1      enp1s0
mlx5_0  1       1       fe80:0000:0000:0000:268a:07ff:feb5:6ef4                 v2      enp1s0
mlx5_0  1       2       0000:0000:0000:0000:0000:ffff:c055:0101 192.85.1.1      v1      enp1s0
mlx5_0  1       3       0000:0000:0000:0000:0000:ffff:c055:0101 192.85.1.1      v2      enp1s0
n_gids_found=4
[root@hsrp1 ~]#
```

Above information shows that INDEX2 is RoCE v1, INDEX3 is RoCE v2.

- Show gids and RoCE version at the client peer

```
[root@mlx-104 ~]# ifconfig enp1s0 192.85.1.2/24 up
[root@mlx-104 ~]# show_gids
DEV     PORT    INDEX   GID                                     IPv4            VER     DEV
---     ----    -----   ---                                     ------------    ---     ---
mlx5_0  1       0       fe80:0000:0000:0000:268a:07ff:feb5:6f54                 v1      enp1s0
mlx5_0  1       1       fe80:0000:0000:0000:268a:07ff:feb5:6f54                 v2      enp1s0
mlx5_0  1       2       0000:0000:0000:0000:0000:ffff:c055:0102 192.85.1.2      v1      enp1s0
mlx5_0  1       3       0000:0000:0000:0000:0000:ffff:c055:0102 192.85.1.2      v2      enp1s0
n_gids_found=4
[root@mlx-104 ~]#
```

Above information shows that INDEX2 is RoCE v1, INDEX3 is RoCE v2

- Test RoCE v1 write latency at server peer

ib_write_lat -a -x 2

- Test RoCE v1 write latency at client peer

ib_write_lat 192.85.1.1 -a --report_gbits -F -x 2

- RoCE v1 write latency value at client peer is shown as below screen shot

```
[root@mlx-104 ~]# ib_write_lat 192.85.1.1 -a --report_gbits -F -x 2
---------------------------------------------------------------------------------------
                    RDMA_Write Latency Test
 Dual-port       : OFF          Device         : mlx5_0
 Number of qps   : 1            Transport type : IB
 Connection type : RC           Using SRQ      : OFF
 TX depth        : 1
 Mtu             : 1024[B]
 Link type       : Ethernet
 GID index       : 2
 Max inline data : 220[B]
 rdma_cm QPs     : OFF
 Data ex. method : Ethernet
---------------------------------------------------------------------------------------
 local address: LID 0000 QPN 0x00d5 PSN 0xcf5bb0 RKey 0x00ae9a VAddr 0x00ffffabf00000
 GID: 00:00:00:00:00:00:00:00:00:00:255:255:192:85:01:02
 remote address: LID 0000 QPN 0x00d8 PSN 0x390d5c RKey 0x0088f4 VAddr 0x00ffff836e0000
 GID: 00:00:00:00:00:00:00:00:00:00:255:255:192:85:01:01
---------------------------------------------------------------------------------------
 #bytes #iterations    t_min[usec]    t_max[usec]  t_typical[usec]    t_avg[usec]    t_stdev[usec]   99% percentile[usec]   99.9% percentile[usec]
 2       1000          1.70           8.15         1.72               1.73           0.03            1.77                   8.15
 4       1000          1.70           2.30         1.72               1.72           0.01            1.75                   2.30
 8       1000          1.70           1.82         1.72               1.73           0.02            1.77                   1.82
 16      1000          1.70           1.82         1.72               1.73           0.02            1.77                   1.82
 32      1000          1.72           2.30         1.75               1.76           0.01            1.80                   2.30
 64      1000          1.75           4.92         1.77               1.78           0.02            1.80                   4.92
 128     1000          1.80           2.62         2.07               1.96           0.13            2.12                   2.62
 256     1000          2.35           2.45         2.40               2.39           0.02            2.42                   2.45
 512     1000          2.50           5.72         2.52               2.53           0.02            2.57                   5.72
 1024    1000          2.67           3.22         2.72               2.72           0.02            2.75                   3.22
 2048    1000          2.87           2.97         2.90               2.91           0.01            2.95                   2.97
 4096    1000          3.22           5.97         3.25               3.26           0.02            3.30                   5.97
 8192    1000          3.87           4.20         3.92               3.92           0.02            3.95                   4.20
 16384   1000          5.25           6.32         5.30               5.34           0.10            5.60                   6.32
 32768   1000          8.80           13.95        9.10               9.11           0.16            9.25                   13.95
 65536   1000          16.77          17.37        17.10              17.00          0.15            17.20                  17.37
 131072  1000          27.27          27.82        27.35              27.38          0.09            27.70                  27.82
 262144  1000          48.27          48.90        48.35              48.38          0.11            48.70                  48.90
 524288  1000          90.24          90.99        90.32              90.40          0.14            90.74                  90.99
 1048576 1000          174.18         175.01       174.53             174.45         0.18            174.81                 175.01
 2097152 1000          342.12         342.87       342.32             342.34         0.11            342.72                 342.87
 4194304 1000          677.97         680.42       678.19             678.22         0.15            678.67                 680.42
 8388608 1000          1349.53        1350.38      1349.81            1349.85        0.16            1350.26                1350.38
---------------------------------------------------------------------------------------
```

- Test RoCE v2 write latency at server peer

ib_write_lat -a -x 3

- Test RoCE v2 write latency at client peer

ib_write_lat 192.85.1.1 -a --report_gbits -F -x 3

- RoCE v2 write latency value at client peer is shown as below screen shot

```
[root@mlx-104 ~]# ib_write_lat 192.85.1.1 -a --report_gbits -F -x 3
---------------------------------------------------------------------------------------
                    RDMA_Write Latency Test
 Dual-port       : OFF          Device         : mlx5_0
 Number of qps   : 1            Transport type : IB
 Connection type : RC           Using SRQ      : OFF
 TX depth        : 1
 Mtu             : 1024[B]
 Link type       : Ethernet
 GID index       : 3
 Max inline data : 220[B]
 rdma_cm QPs     : OFF
 Data ex. method : Ethernet
---------------------------------------------------------------------------------------
 local address: LID 0000 QPN 0x00d6 PSN 0xfff946 RKey 0x00c1ad VAddr 0x00ffff81a20000
 GID: 00:00:00:00:00:00:00:00:00:00:255:255:192:85:01:02
 remote address: LID 0000 QPN 0x00d9 PSN 0x79dcd RKey 0x0078e3 VAddr 0x00ffff81b80000
 GID: 00:00:00:00:00:00:00:00:00:00:255:255:192:85:01:01
---------------------------------------------------------------------------------------
 #bytes #iterations    t_min[usec]    t_max[usec]  t_typical[usec]    t_avg[usec]    t_stdev[usec]   99% percentile[usec]   99.9% percentile[usec]
 2       1000          1.67           2.67         1.72               1.71           0.02            1.75                   2.67
 4       1000          1.67           6.05         1.70               1.71           0.02            1.75                   6.05
 8       1000          1.67           2.20         1.72               1.71           0.01            1.75                   2.20
 16      1000          1.70           1.77         1.72               1.72           0.01            1.75                   1.77
 32      1000          1.72           1.82         1.75               1.75           0.01            1.77                   1.82
 64      1000          1.72           2.30         1.75               1.76           0.01            1.80                   2.30
 128     1000          1.80           5.72         2.05               1.95           0.13            2.10                   5.72
 256     1000          2.32           2.90         2.37               2.37           0.01            2.40                   2.90
 512     1000          2.50           2.57         2.52               2.52           0.01            2.55                   2.57
 1024    1000          2.67           6.85         2.72               2.72           0.02            2.75                   6.85
 2048    1000          2.87           2.95         2.90               2.91           0.01            2.95                   2.95
 4096    1000          3.20           6.12         3.25               3.24           0.02            3.27                   6.12
 8192    1000          3.87           4.20         3.90               3.91           0.02            3.95                   4.20
 16384   1000          5.22           7.40         5.27               5.31           0.10            5.57                   7.40
 32768   1000          8.87           13.20        9.05               9.07           0.07            9.22                   13.20
 65536   1000          16.45          17.02        16.77              16.68          0.15            16.97                  17.02
 131072  1000          26.85          27.47        26.92              26.95          0.09            27.25                  27.47
 262144  1000          47.60          48.20        47.67              47.72          0.11            48.05                  48.20
 524288  1000          89.12          89.74        89.22              89.28          0.13            89.59                  89.74
 1048576 1000          172.19         173.11       172.41             172.45         0.20            172.86                 173.11
 2097152 1000          338.30         339.17       338.77             338.75         0.14            339.00                 339.17
 4194304 1000          670.46         671.84       670.76             671.02         0.45            671.74                 671.84
 8388608 1000          1334.85        1337.45      1335.13            1335.35        0.42            1336.15                1337.45
---------------------------------------------------------------------------------------
[root@mlx-104 ~]#
```

- Test RoCE v1 write bandwidth at server peer

ib_write_bw -a -x 2

- Test RoCE v1 write bandwidth at client peer

ib_write_bw 192.85.1.1 -a --report_gbits -F -x 2

- RoCE v1 write bandwidth value at client peer is shown as below screen shot

```
[root@mlx-104 ~]# ib_write_bw 192.85.1.1 -a --report_gbits -F -x 2
---------------------------------------------------------------------------------------
                    RDMA_Write BW Test
 Dual-port       : OFF          Device         : mlx5_0
 Number of qps   : 1            Transport type : IB
 Connection type : RC           Using SRQ      : OFF
 TX depth        : 128
 CQ Moderation   : 100
 Mtu             : 1024[B]
 Link type       : Ethernet
 GID index       : 2
 Max inline data : 0[B]
 rdma_cm QPs     : OFF
 Data ex. method : Ethernet
---------------------------------------------------------------------------------------
 local address: LID 0000 QPN 0x00d8 PSN 0xff208c RKey 0x00ccb6 VAddr 0x00ffffaa340000
 GID: 00:00:00:00:00:00:00:00:00:00:255:255:192:85:01:02
 remote address: LID 0000 QPN 0x00db PSN 0xe68838 RKey 0x00ae98 VAddr 0x00ffff933b0000
 GID: 00:00:00:00:00:00:00:00:00:00:255:255:192:85:01:01
---------------------------------------------------------------------------------------
 #bytes     #iterations    BW peak[Gb/sec]    BW average[Gb/sec]   MsgRate[Mpps]
 2          5000           0.029092            0.026448            1.653008
 4          5000           0.091436            0.086568            2.705254
 8          5000             0.18               0.17               2.704988
 16         5000             0.37               0.35               2.713548
 32         5000             0.73               0.69               2.711137
 64         5000             1.46               1.39               2.715009
 128        5000             2.93               2.77               2.709133
 256        5000             5.85               5.54               2.703500
 512        5000             11.70              10.43              2.547520
 1024       5000             20.48              20.12              2.456182
 2048       5000             36.41              36.05              2.200291
 4096       5000             50.42              49.54              1.511975
 8192       5000             50.42              49.80              0.759834
 16384      5000             50.42              49.88              0.380558
 32768      5000             49.94              49.92              0.190432
 65536      5000             50.17              49.94              0.095254
 131072     5000             50.06              49.95              0.047637
 262144     5000             50.00              49.95              0.023820
 524288     5000             49.97              49.96              0.011911
 1048576    5000             49.97              49.96              0.005956
 2097152    5000             49.97              49.96              0.002978
 4194304    5000             49.96              49.96              0.001489
 8388608    5000             49.96              49.96              0.000744
---------------------------------------------------------------------------------------
```

- Test RoCE v2 write bandwidth at server peer

ib_write_bw -a -x 3

- Test RoCE v2 write bandwidth at client peer

ib_write_bw 192.85.1.1 -a --report_gbits -F -x 3

- RoCE v2 write bandwidth value at client peer is shown as below screen shot

```
[root@mlx-104 ~]# ib_write_bw 192.85.1.1 -a --report_gbits -F -x 3
---------------------------------------------------------------------------------------
                    RDMA_Write BW Test
 Dual-port       : OFF          Device         : mlx5_0
 Number of qps   : 1            Transport type : IB
 Connection type : RC           Using SRQ      : OFF
 TX depth        : 128
 CQ Moderation   : 100
 Mtu             : 1024[B]
 Link type       : Ethernet
 GID index       : 3
 Max inline data : 0[B]
 rdma_cm QPs     : OFF
 Data ex. method : Ethernet
---------------------------------------------------------------------------------------
 local address: LID 0000 QPN 0x00d9 PSN 0xc782ac RKey 0x00e6d1 VAddr 0x00ffff906c0000
 GID: 00:00:00:00:00:00:00:00:00:00:255:255:192:85:01:02
 remote address: LID 0000 QPN 0x00dc PSN 0x5684bd RKey 0x00bda7 VAddr 0x00ffffab8f0000
 GID: 00:00:00:00:00:00:00:00:00:00:255:255:192:85:01:01
---------------------------------------------------------------------------------------
 #bytes     #iterations    BW peak[Gb/sec]    BW average[Gb/sec]   MsgRate[Mpps]
 2          5000           0.024617            0.023742            1.483865
 4          5000           0.080006            0.072674            2.271048
 8          5000             0.16               0.15              2.275528
 16         5000             0.32               0.29              2.275486
 32         5000             0.64               0.58              2.276917
 64         5000             1.28               1.17              2.276751
 128        5000             2.56               2.32              2.264133
 256        5000             5.12               4.59              2.238951
 512        5000             9.10               8.99              2.194695
 1024       5000             18.21              17.41             2.125477
 2048       5000             32.77              32.36             1.975352
 4096       5000             50.42              50.12             1.529693
 8192       5000             50.42              50.34             0.768094
 16384      5000             51.41              50.42             0.384690
 32768      5000             50.91              50.46             0.192498
 65536      5000             50.66              50.48             0.096288
 131072     5000             50.54              50.49             0.048153
 262144     5000             50.54              50.50             0.024079
 524288     5000             50.51              50.50             0.012040
 1048576    5000             50.51              50.50             0.006020
 2097152    5000             50.51              50.50             0.003010
 4194304    5000             50.50              50.50             0.001505
 8388608    5000             50.50              50.50             0.000753
---------------------------------------------------------------------------------------
```

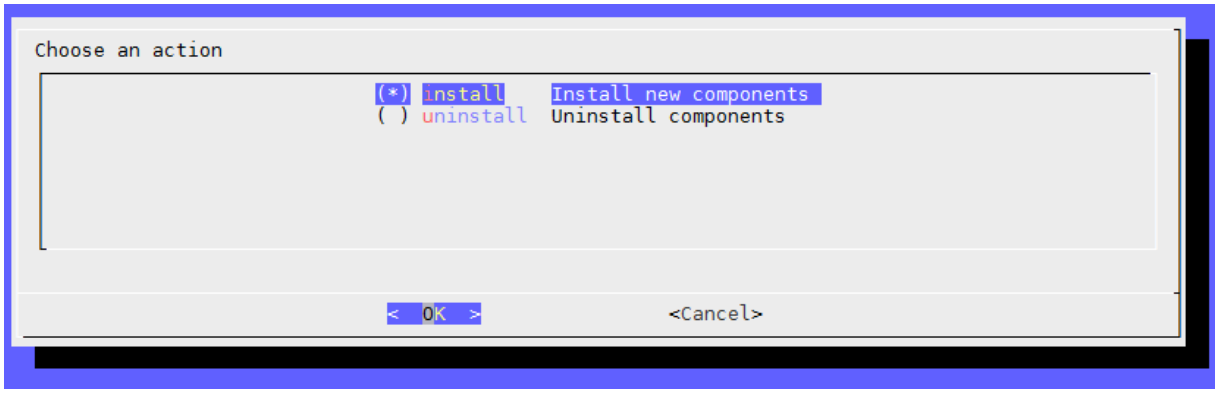
4.3 iWARP

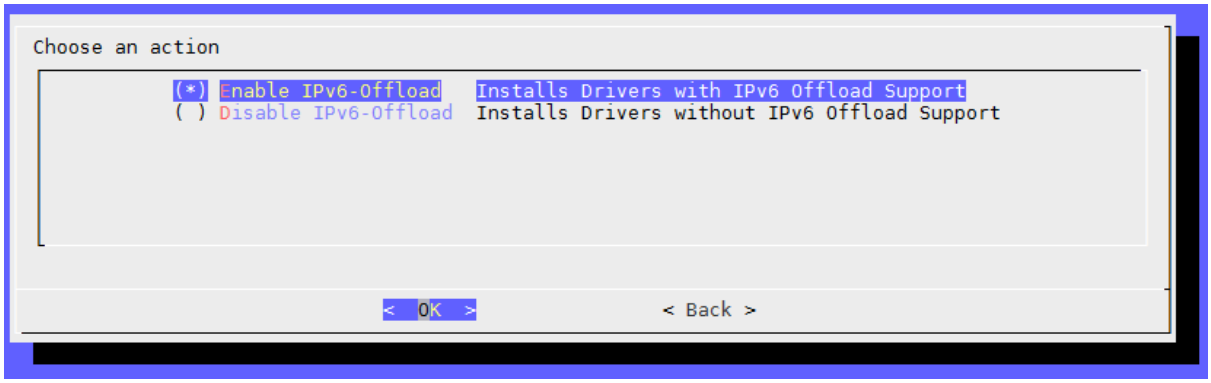
4.3.1 Install kernel driver

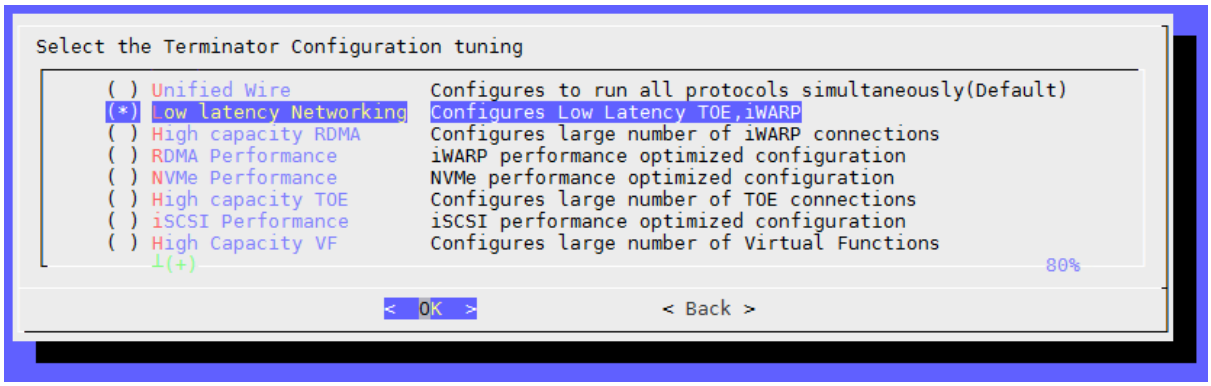
Download ChelsioUwire-3.8.0.2.tar.gz from Chelsio official website, then untar the package to a tmp folder such as /tmp/, change into the folder, then run the installation scripts:

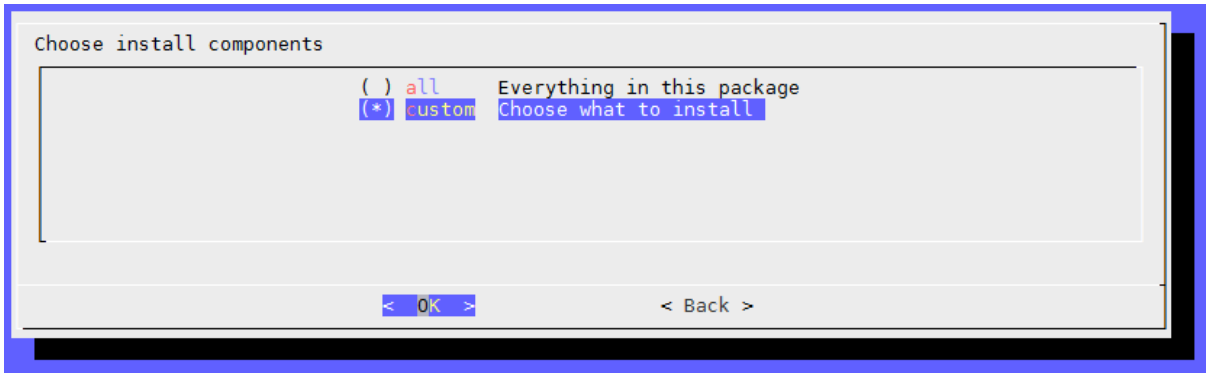
./install.py

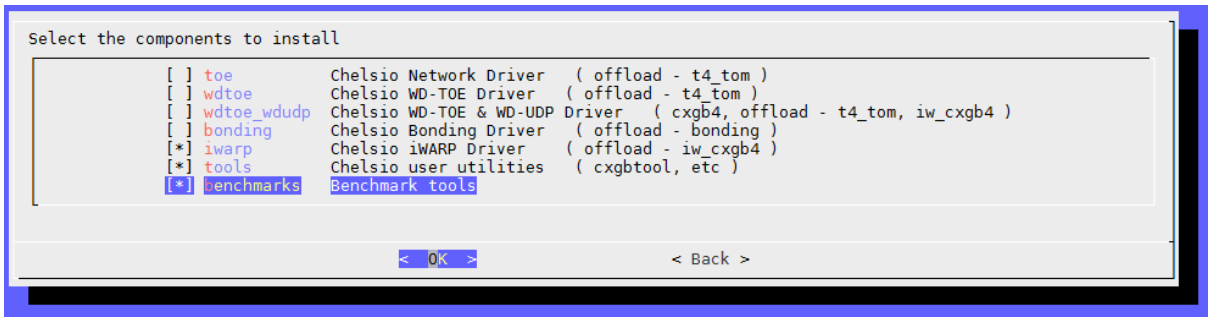
Then press OK button according to below screen shot.

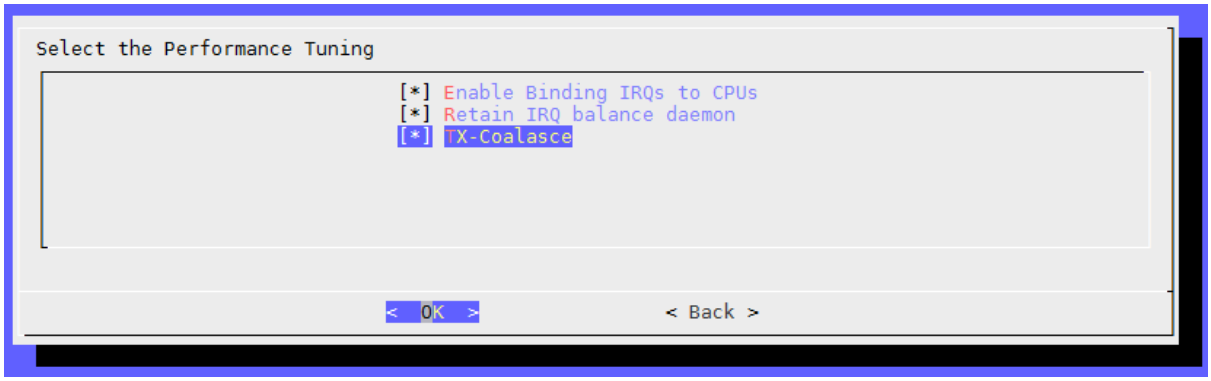

Choose an action
(*) install Install new components
( ) uninstall Uninstall components
< OK >
<Cancel>

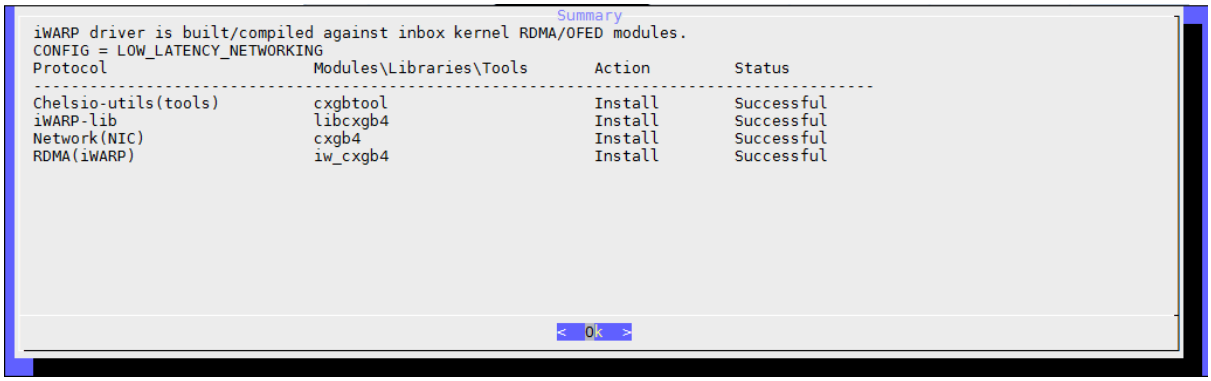

Choose an action
(*) Enable IPv6-Offload Installs Drivers with IPv6 Offload Support
( ) Disable IPv6-Offload Installs Drivers without IPv6 Offload Support
< OK >
< Back >


Select the Terminator Configuration tuning
( ) Unified Wire Configures to run all protocols simultaneously(Default)
(*) Low latency Networking Configures Low Latency TOE,iWARP
( ) High capacity RDMA Configures large number of iWARP connections
( ) RDMA Performance iWARP performance optimized configuration
( ) NVMe Performance NVMe performance optimized configuration
( ) High capacity TOE Configures large number of TOE connections
( ) iSCSI Performance iSCSI performance optimized configuration
( ) High Capacity VF Configures large number of Virtual Functions
80%
< OK >
< Back >


Choose install components
( ) all Everything in this package
(*) custom Choose what to install
< OK >
< Back >







4.3.2 Install user space tool and libraries

#yum install -y automake bc elfutils-libelf-devel epel-release gcc gcc-c++ libibverbs libibverbs-devel libibverbs-utils libnl-devel libnl3-devel librdmacm librdmacm-devel librdmacm-utils ncurses-devel openssl-devel perftest perl-Switch valgrind-devel

4.3.3 Test steps

4.3.3.1 Common setting

1. Configure server NIC with an IPv4 address:

ifconfig enP5p1s0f4 192.85.1.1/24 up

2. Configure client NIC with an IPv4 address:

ifconfig enP5p1s0f4 192.85.1.2/24 up

4.3.3.2 Test latency

- Test iWARP write latency at the server peer:

ib_write_lat -d cxgb4_0 -i 1 -R –a

- Test iWARP write latency at the client peer:

ib_write_lat 192.85.1.1 -a -d cxgb4_0 -i 1 –R

- Server peer shows the NIC write latency result in iwarp as below:

```
[root@chelsio-102 ChelsioUwire-3.8.0.2]# ib_write_lat -d cxgb4_0 -i 1 -R -a

************************************
* Waiting for client to connect... *
************************************
---------------------------------------------------------------------------------------
                    RDMA_Write Latency Test
 Dual-port       : OFF          Device         : cxgb4_0
 Number of qps   : 1            Transport type : IW
 Connection type : RC           Using SRQ      : OFF
 Mtu             : 1024[B]
 Link type       : Ethernet
 GID index       : 0
 Max inline data : 220[B]
 rdma_cm QPs     : ON
 Data ex. method : rdma_cm
---------------------------------------------------------------------------------------
 Waiting for client rdma_cm QP to connect
 Please run the same command with the IB/RoCE interface IP
---------------------------------------------------------------------------------------
 local address: LID 0000 QPN 0x0582 PSN 0xda736a
 GID: 00:07:67:62:138:240:00:00:00:00:00:00:00:00:00:00
 remote address: LID 0000 QPN 0x0582 PSN 0x723f56
 GID: 00:07:67:62:139:80:00:00:00:00:00:00:00:00:00:00
---------------------------------------------------------------------------------------
 #bytes #iterations    t_min[usec]    t_max[usec]  t_typical[usec]    t_avg[usec]    t_stdev[usec]   99% percentile[usec]   99.9% percentile[usec]
 2       1000          3.40           11.15        3.45               3.45           0.08            3.57                   11.15
 4       1000          3.37           5.10         3.42               3.44           0.05            3.55                   5.10
 8       1000          3.37           5.30         3.42               3.44           0.06            3.55                   5.30
 16      1000          3.42           5.22         3.47               3.47           0.05            3.57                   5.22
 32      1000          3.47           5.25         3.52               3.52           0.05            3.62                   5.25
 64      1000          3.52           3.70         3.55               3.56           0.03            3.65                   3.70
 128     1000          3.65           6.37         3.70               3.72           0.14            4.30                   6.37
 256     1000          4.55           7.20         4.77               4.80           0.12            5.10                   7.20
 512     1000          4.82           8.20         5.30               5.31           0.08            5.45                   8.20
 1024    1000          5.37           7.02         5.60               5.61           0.07            5.75                   7.02
 2048    1000          5.75           10.97        5.92               5.93           0.07            6.07                   10.97
 4096    1000          6.40           7.67         6.45               6.46           0.06            6.57                   7.67
 8192    1000          7.75           8.55         7.82               7.81           0.02            7.85                   8.55
 16384   1000          10.62          12.00        10.65              10.67          0.06            10.77                  12.00
 32768   1000          16.30          17.22        16.35              16.35          0.04            16.47                  17.22
 65536   1000          27.52          28.45        27.57              27.57          0.04            27.70                  28.45
 131072  1000          49.95          50.87        50.00              50.01          0.05            50.12                  50.87
 262144  1000          94.79          95.07        94.84              94.85          0.05            95.02                  95.07
 524288  1000          184.61         185.41       184.66             184.68         0.04            184.81                 185.41
 1048576 1000          364.07         366.47       364.12             364.14         0.05            364.27                 366.47
 2097152 1000          723.13         724.06       723.21             723.21         0.04            723.33                 724.06
 4194304 1000          1441.31        1442.91      1441.38            1441.40        0.05            1441.53                1442.91
 8388608 1000          2877.47        2877.79      2877.57            2877.58        0.04            2877.72                2877.79
---------------------------------------------------------------------------------------
[root@chelsio-102 ChelsioUwire-3.8.0.2]#
```

- Client peer shows the NIC write latency result in iwarp as below:

```
[root@chelsio-104 ChelsioUwire-3.8.0.2]# ib_write_lat 192.85.1.1 -a -d cxgb4_0 -i 1 -R
---------------------------------------------------------------------------------
                    RDMA_Write Latency Test
 Dual-port       : OFF          Device         : cxgb4_0
 Number of qps   : 1            Transport type : IW
 Connection type : RC           Using SRQ      : OFF
 TX depth        : 1
 Mtu             : 1024[B]
 Link type       : Ethernet
 GID index       : 0
 Max inline data : 220[B]
 rdma_cm QPs     : ON
 Data ex. method : rdma_cm
---------------------------------------------------------------------------------
 local address: LID 0000 QPN 0x0582 PSN 0x723f56
 GID: 00:07:67:62:139:80:00:00:00:00:00:00:00:00:00:00
 remote address: LID 0000 QPN 0x0582 PSN 0xda736a
 GID: 00:07:67:62:138:240:00:00:00:00:00:00:00:00:00:00
---------------------------------------------------------------------------------
 #bytes #iterations    t_min[usec]    t_max[usec]  t_typical[usec]    t_avg[usec]    t_stdev[usec]   99% percentile[usec]   99.9% percentile[usec]
 2       1000          3.40           10.92        3.45               3.45           0.07            3.57                   10.92
 4       1000          3.40           5.12         3.42               3.44           0.03            3.55                   5.12
 8       1000          3.37           5.47         3.42               3.44           0.03            3.55                   5.47
 16      1000          3.42           4.65         3.47               3.47           0.05            3.57                   4.65
 32      1000          3.47           5.30         3.52               3.52           0.05            3.62                   5.30
 64      1000          3.52           3.72         3.55               3.56           0.02            3.65                   3.72
 128     1000          3.65           6.15         3.70               3.72           0.10            4.05                   6.15
 256     1000          4.52           5.52         4.80               4.80           0.12            5.10                   5.52
 512     1000          4.82           8.25         5.30               5.31           0.08            5.45                   8.25
 1024    1000          5.32           7.02         5.60               5.61           0.08            5.77                   7.02
 2048    1000          5.72           11.07        5.92               5.93           0.06            6.07                   11.07
 4096    1000          6.37           7.62         6.45               6.46           0.05            6.57                   7.62
 8192    1000          7.75           8.22         7.82               7.81           0.02            7.85                   8.22
 16384   1000          10.62          12.12        10.65              10.67          0.05            10.77                  12.12
 32768   1000          16.30          17.17        16.35              16.35          0.04            16.45                  17.17
 65536   1000          27.52          28.22        27.57              27.57          0.04            27.67                  28.22
 131072  1000          49.95          50.82        50.00              50.01          0.05            50.17                  50.82
 262144  1000          94.79          95.04        94.84              94.85          0.04            95.02                  95.04
 524288  1000          184.58         185.18       184.68             184.68         0.04            184.81                 185.18
 1048576 1000          364.07         366.52       364.12             364.14         0.04            364.25                 366.52
 2097152 1000          723.15         723.70       723.22             723.22         0.04            723.37                 723.70
 4194304 1000          1441.31        1442.94      1441.39            1441.40        0.05            1441.54                1442.94
 8388608 1000          2877.45        2877.75      2877.55            2877.56        0.05            2877.70                2877.75
---------------------------------------------------------------------------------
[root@chelsio-104 ChelsioUwire-3.8.0.2]#
```

- Test read latency at server peer:

ib_read_lat -d cxgb4_0 -i 1 -R -a

- Test read latency at client peer:

ib_read_lat 192.85.1.1 -a -d cxgb4_0 -i 1 –R

- Client read latency screen shot:

```
[root@chelsio-104 ChelsioUwire-3.8.0.2]# ib_read_lat 192.85.1.1 -a -d cxgb4_0 -i 1 -R
---------------------------------------------------------------------------------------
                    RDMA_Read Latency Test
 Dual-port       : OFF          Device         : cxgb4_0
 Number of qps   : 1            Transport type : IW
 Connection type : RC           Using SRQ      : OFF
 TX depth        : 1
 Mtu             : 1024[B]
 Link type       : Ethernet
 GID index       : 0
 Outstand reads  : 21
 rdma_cm QPs     : ON
 Data ex. method : rdma_cm
---------------------------------------------------------------------------------------
 local address: LID 0000 QPN 0x0602 PSN 0xe358c0
 GID: 00:07:67:62:139:80:00:00:00:00:00:00:00:00:00:00
 remote address: LID 0000 QPN 0x0602 PSN 0xa85283
 GID: 00:07:67:62:138:240:00:00:00:00:00:00:00:00:00:00
---------------------------------------------------------------------------------------
 #bytes #iterations    t_min[usec]    t_max[usec]  t_typical[usec]    t_avg[usec]    t_stdev[usec]   99% percentile[usec]   99.9% percentile[usec]
 2       1000          6.20           7.30         6.35               6.37           0.07            6.60                   7.30
 4       1000          6.20           7.80         6.35               6.36           0.07            6.60                   7.80
 8       1000          6.20           6.70         6.35               6.35           0.06            6.60                   6.70
 16      1000          6.20           7.50         6.35               6.37           0.08            6.60                   7.50
 32      1000          6.25           6.85         6.35               6.39           0.09            6.75                   6.85
 64      1000          6.30           9.30         6.40               6.42           0.08            6.70                   9.30
 128     1000          6.35           7.45         6.50               6.50           0.07            6.75                   7.45
 256     1000          6.50           9.45         6.65               6.66           0.09            6.90                   9.45
 512     1000          6.75           7.25         6.90               6.92           0.07            7.15                   7.25
 1024    1000          7.20           8.50         7.35               7.39           0.09            7.75                   8.50
 2048    1000          7.70           9.45         7.90               7.92           0.10            8.30                   9.45
 4096    1000          8.50           10.15        8.70               8.70           0.08            8.95                   10.15
 8192    1000          9.80           11.15        10.00              10.01          0.07            10.25                  11.15
 16384   1000          12.65          13.95        12.75              12.78          0.07            13.00                  13.95
 32768   1000          18.15          19.60        18.45              18.45          0.07            18.70                  19.60
 65536   1000          29.45          30.75        29.65              29.67          0.09            29.95                  30.75
 131072  1000          51.90          53.30        52.10              52.10          0.07            52.25                  53.30
 262144  1000          96.69          97.24        96.94              96.94          0.07            97.14                  97.24
 524288  1000          186.48         187.88       186.78             186.77         0.08            186.98                 187.88
 1048576 1000          366.02         367.47       366.22             366.24         0.07            366.47                 367.47
 2097152 1000          725.04         725.74       725.29             725.30         0.07            725.49                 725.74
 4194304 1000          1443.12        1444.52      1443.42            1443.45        0.08            1443.67                1444.52
 8388608 1000          2879.34        2879.94      2879.64            2879.63        0.07            2879.84                2879.94
---------------------------------------------------------------------------------------
[root@chelsio-104 ChelsioUwire-3.8.0.2]#
```

- Test send latency at server peer:

ib_send_lat -d cxgb4_0 -i 1 -R –a

- Test send latency at client peer:

ib_send_lat 192.85.1.1 -a -d cxgb4_0 -i 1 –R

- Client send latency screen shot:

```
[root@chelsio-104 ChelsioUwire-3.8.0.2]# ib_send_lat 192.85.1.1 -a -d cxgb4_0 -i 1 -R
---------------------------------------------------------------------------------------
                    Send Latency Test
 Dual-port       : OFF          Device         : cxgb4_0
 Number of qps   : 1            Transport type : IW
 Connection type : RC           Using SRQ      : OFF
 TX depth        : 1
 Mtu             : 1024[B]
 Link type       : Ethernet
 GID index       : 0
 Max inline data : 236[B]
 rdma_cm QPs     : ON
 Data ex. method : rdma_cm
---------------------------------------------------------------------------------------
 local address: LID 0000 QPN 0x0682 PSN 0x37ae0e
 GID: 00:07:67:62:139:80:00:00:00:00:00:00:00:00:00:00
 remote address: LID 0000 QPN 0x0682 PSN 0xcd2ee4
 GID: 00:07:67:62:138:240:00:00:00:00:00:00:00:00:00:00
---------------------------------------------------------------------------------------
 #bytes #iterations    t_min[usec]    t_max[usec]  t_typical[usec]    t_avg[usec]    t_stdev[usec]   99% percentile[usec]   99.9% percentile[usec]
 2       1000          3.65           6.57         3.85               3.87           0.16            4.22                   6.57
 4       1000          3.62           6.55         3.80               3.83           0.14            4.07                   6.55
 8       1000          3.62           6.00         3.80               3.82           0.14            4.05                   6.00
 16      1000          3.65           10.12        3.85               3.86           0.14            4.12                   10.12
 32      1000          3.70           6.45         3.90               3.91           0.15            4.17                   6.45
 64      1000          3.75           10.20        3.95               3.98           0.14            4.22                   10.20
 128     1000          3.87           7.40         4.12               4.12           0.22            5.05                   7.40
 256     1000          4.82           8.50         5.27               5.35           0.21            5.80                   8.50
 512     1000          5.30           7.82         5.57               5.61           0.15            5.92                   7.82
 1024    1000          5.65           10.85        5.87               5.92           0.16            6.25                   10.85
 2048    1000          6.00           9.52         6.27               6.31           0.18            6.62                   9.52
 4096    1000          6.62           9.40         6.80               6.83           0.16            7.07                   9.40
 8192    1000          8.00           10.10        8.15               8.18           0.13            8.42                   10.10
 16384   1000          10.85          12.20        11.02              11.07          0.14            11.35                  12.20
 32768   1000          16.57          17.70        16.77              16.78          0.15            17.02                  17.70
 65536   1000          27.80          28.95        27.97              28.01          0.13            28.25                  28.95
 131072  1000          50.30          51.80        50.50              50.52          0.17            50.85                  51.80
 262144  1000          95.32          96.37        95.49              95.52          0.14            95.82                  96.37
 524288  1000          185.29         186.69       185.51             185.51         0.13            185.71                 186.69
 1048576 1000          365.29         366.34       365.47             365.49         0.13            365.72                 366.34
 2097152 1000          725.39         725.92       725.64             725.63         0.13            725.87                 725.92
 4194304 1000          1445.51        1446.71      1445.71            1445.74        0.15            1445.96                1446.71
 8388608 1000          2885.85        2886.98      2886.08            2886.07        0.14            2886.35                2886.98
---------------------------------------------------------------------------------------
[root@chelsio-104 ChelsioUwire-3.8.0.2]#
```

4.3.3.3 Test uni-directional bandwidth

- Test write bandwidth at server peer:

ib_write_bw -d cxgb4_0 -i 1 -R -a -F --report_gbits

- Test write bandwidth at client peer:

ib_write_bw 192.85.1.1 -a -d cxgb4_0 -i 1 -R -F --report_gbits

- Client write bandwidth screen shot:

```
[root@chelsio-104 ChelsioUwire-3.8.0.2]# ib_write_bw 192.85.1.1 -a -d cxgb4_0 -i 1 -R -F --report_gbits
---------------------------------------------------------------------------------------
                    RDMA_Write BW Test
 Dual-port       : OFF          Device         : cxgb4_0
 Number of qps   : 1            Transport type : IW
 Connection type : RC           Using SRQ      : OFF
 TX depth        : 128
 CQ Moderation   : 100
 Mtu             : 1024[B]
 Link type       : Ethernet
 GID index       : 0
 Max inline data : 0[B]
 rdma_cm QPs     : ON
 Data ex. method : rdma_cm
---------------------------------------------------------------------------------------
 local address: LID 0000 QPN 0x0702 PSN 0x99763c
 GID: 00:07:67:62:139:80:00:00:00:00:00:00:00:00:00:00
 remote address: LID 0000 QPN 0x0702 PSN 0x124879
 GID: 00:07:67:62:138:240:00:00:00:00:00:00:00:00:00:00
---------------------------------------------------------------------------------------
 #bytes     #iterations    BW peak[Gb/sec]    BW average[Gb/sec]   MsgRate[Mpps]
 2          5000           0.032002            0.030050            1.878142
 4          5000             0.13               0.11              3.593148
 8          5000             0.26               0.23              3.603470
 16         5000             0.51               0.46              3.603111
 32         5000             1.02               0.92              3.604762
 64         5000             2.05               1.84              3.600102
 128        5000             4.10               3.66              3.578637
 256        5000             8.19               7.28              3.553803
 512        5000             16.39              14.27             3.484020
 1024       5000             23.41              22.82             2.786193
 2048       5000             23.41              22.78             1.390359
 4096       5000             24.27              23.31             0.711401
 8192       5000             23.83              23.30             0.355485
 16384      5000             23.41              23.29             0.177723
 32768      5000             23.51              23.36             0.089102
 65536      5000             23.41              23.36             0.044554
 131072     5000             23.38              23.36             0.022277
 262144     5000             23.37              23.36             0.011139
 524288     5000             23.37              23.36             0.005569
 1048576    5000             23.37              23.36             0.002785
 2097152    5000             23.36              23.36             0.001392
 4194304    5000             23.36              23.36             0.000696
 8388608    5000             23.36              23.36             0.000348
---------------------------------------------------------------------------------------
[root@chelsio-104 ChelsioUwire-3.8.0.2]#
```

- Test read bandwidth at server peer:

ib_read_bw -d cxgb4_0 -i 1 -R -a -F --report_gbits

- Test read bandwidth at client peer:

ib_read_bw 192.85.1.1 -a -d cxgb4_0 -i 1 -R -F --report_gbits

- Client read bandwidth screen shot:

```
[root@chelsio-104 ChelsioUwire-3.8.0.2]# ib_read_bw 192.85.1.1 -a -d cxgb4_0 -i 1 -R -F --report_gbits
---------------------------------------------------------------------------------------
                    RDMA_Read BW Test
 Dual-port       : OFF          Device         : cxgb4_0
 Number of qps   : 1            Transport type : IW
 Connection type : RC           Using SRQ      : OFF
 TX depth        : 128
 CQ Moderation   : 100
 Mtu             : 1024[B]
 Link type       : Ethernet
 GID index       : 0
 Outstand reads  : 21
 rdma_cm QPs     : ON
 Data ex. method : rdma_cm
---------------------------------------------------------------------------------------
 local address: LID 0000 QPN 0x0782 PSN 0x3b44e2
 GID: 00:07:67:62:139:80:00:00:00:00:00:00:00:00:00:00
 remote address: LID 0000 QPN 0x0782 PSN 0x9e06
 GID: 00:07:67:62:138:240:00:00:00:00:00:00:00:00:00:00
---------------------------------------------------------------------------------------
 #bytes     #iterations    BW peak[Gb/sec]    BW average[Gb/sec]   MsgRate[Mpps]
 2          1000           0.024617           0.021760             1.359998
 4          1000           0.091436           0.083329             2.604028
 8          1000             0.18               0.17               2.715755
 16         1000             0.37               0.35               2.716511
 32         1000             0.73               0.69               2.712803
 64         1000             1.46               1.39               2.720167
 128        1000             2.93               2.78               2.711346
 256        1000             5.85               5.56               2.717227
 512        1000             11.70              10.74              2.621440
 1024       1000             20.48              19.58              2.390514
 2048       1000             23.41              21.99              1.342208
 4096       1000             23.41              22.72              0.693414
 8192       1000             23.41              23.05              0.351737
 16384      1000             23.41              23.21              0.177088
 32768      1000             23.30              23.28              0.088817
 65536      1000             23.36              23.32              0.044485
 131072     1000             23.36              23.34              0.022259
 262144     1000             23.36              23.35              0.011134
 524288     1000             23.36              23.36              0.005568
 1048576    1000             23.36              23.36              0.002785
 2097152    1000             23.36              23.36              0.001392
 4194304    1000             23.36              23.36              0.000696
 8388608    1000             23.36              23.36              0.000348
---------------------------------------------------------------------------------------
[root@chelsio-104 ChelsioUwire-3.8.0.2]#
```

- Test send bandwidth at server peer:

ib_send_bw -d cxgb4_0 -i 1 -R -a -F --report_gbits

- Test send bandwidth at client peer:

ib_send_bw 192.85.1.1 -a -d cxgb4_0 -i 1 -R -F --report_gbits

- Client send bandwidth screen shot:

```
[root@chelsio-104 ChelsioUwire-3.8.0.2]# ib_send_bw 192.85.1.1 -a -d cxgb4_0 -i 1 -R -F --report_gbits
---------------------------------------------------------------------------------------
                    Send BW Test
 Dual-port       : OFF          Device         : cxgb4_0
 Number of qps   : 1            Transport type : IW
 Connection type : RC           Using SRQ      : OFF
 TX depth        : 128
 CQ Moderation   : 100
 Mtu             : 1024[B]
 Link type       : Ethernet
 GID index       : 0
 Max inline data : 0[B]
 rdma_cm QPs     : ON
 Data ex. method : rdma_cm
---------------------------------------------------------------------------------------
 local address: LID 0000 QPN 0x0802 PSN 0xf629cd
 GID: 00:07:67:62:139:80:00:00:00:00:00:00:00:00:00:00
 remote address: LID 0000 QPN 0x0802 PSN 0xfa03f4
 GID: 00:07:67:62:138:240:00:00:00:00:00:00:00:00:00:00
---------------------------------------------------------------------------------------
 #bytes     #iterations    BW peak[Gb/sec]    BW average[Gb/sec]   MsgRate[Mpps]
 2          1000           0.035558            0.031252            1.953242
 4          1000             0.13               0.11              3.427280
 8          1000             0.26               0.22              3.462855
 16         1000             0.51               0.44              3.468882
 32         1000             1.02               0.89              3.472489
 64         1000             2.05               1.77              3.462299
 128        1000             4.10               3.55              3.469469
 256        1000             8.19               7.07              3.450928
 512        1000             16.39              13.82             3.375216
 1024       1000             23.41              22.77             2.779519
 2048       1000             23.41              22.73             1.387447
 4096       1000             24.27              23.28             0.710433
 8192       1000             23.83              23.25             0.354815
 16384      1000             23.41              23.24             0.177308
 32768      1000             23.30              23.24             0.088637
 65536      1000             23.30              23.27             0.044378
 131072     1000             23.30              23.28             0.022204
 262144     1000             23.30              23.29             0.011106
 524288     1000             23.30              23.29             0.005554
 1048576    1000             23.30              23.30             0.002777
 2097152    1000             23.30              23.30             0.001389
 4194304    1000             23.30              23.30             0.000694
 8388608    1000             23.30              23.30             0.000347
---------------------------------------------------------------------------------------
[root@chelsio-104 ChelsioUwire-3.8.0.2]#
```

4.3.3.4 Test uni-directional bandwidth

- Test bi-directional write bandwidth at server peer:

ib_write_bw -d cxgb4_0 -i 1 -R -a -F --report_gbits -b

- Test bi-directional write bandwidth at client peer:

ib_write_bw 192.85.1.1 -a -d cxgb4_0 -i 1 -R -F --report_gbits -b

- Client bi-directional write bandwidth screen shot:

```
[root@chelsio-104 ChelsioUwire-3.8.0.2]# ib_write_bw 192.85.1.1 -a -d cxgb4_0 -i 1 -R -F --report_gbits -b
---------------------------------------------------------------------------------------
                    RDMA_Write Bidirectional BW Test
 Dual-port       : OFF          Device         : cxgb4_0
 Number of qps   : 1            Transport type : IW
 Connection type : RC           Using SRQ      : OFF
 TX depth        : 128
 CQ Moderation   : 100
 Mtu             : 1024[B]
 Link type       : Ethernet
 GID index       : 0
 Max inline data : 0[B]
 rdma_cm QPs     : ON
 Data ex. method : rdma_cm
---------------------------------------------------------------------------------------
 local address: LID 0000 QPN 0x0882 PSN 0x1c404d
 GID: 00:07:67:62:139:80:00:00:00:00:00:00:00:00:00:00
 remote address: LID 0000 QPN 0x0882 PSN 0x794f6a
 GID: 00:07:67:62:138:240:00:00:00:00:00:00:00:00:00:00
---------------------------------------------------------------------------------------
 #bytes     #iterations    BW peak[Gb/sec]    BW average[Gb/sec]   MsgRate[Mpps]
 2          5000           0.067559            0.062267            3.891710
 4          5000             0.26               0.23              7.115060
 8          5000             0.51               0.46              7.115288
 16         5000             1.02               0.91              7.122918
 32         5000             2.05               1.82              7.125684
 64         5000             4.10               3.64              7.114803
 128        5000             8.19               7.23              7.061257
 256        5000             16.39              14.38             7.022824
 512        5000             32.77              28.19             6.881792
 1024       5000             46.82              44.07             5.380175
 2048       5000             46.82              44.74             2.730881
 4096       5000             46.81              45.63             1.392655
 8192       5000             46.40              45.64             0.696443
 16384      5000             45.59              45.45             0.346787
 32768      5000             46.30              46.12             0.175936
 65536      5000             46.20              46.10             0.087927
 131072     5000             46.20              46.16             0.044018
 262144     5000             46.18              46.17             0.022014
 524288     5000             46.18              46.16             0.011006
 1048576    5000             46.16              46.15             0.005502
 2097152    5000             46.15              46.15             0.002751
 4194304    5000             46.15              46.15             0.001375
 8388608    5000             46.16              46.16             0.000688
---------------------------------------------------------------------------------------
[root@chelsio-104 ChelsioUwire-3.8.0.2]#
```

5. Reference links

- Remote direct memory access wikipedia
- HowTo Configure Soft-RoCE
- How to configure Soft-RoCE with Mellanox OFED 4.2
- HowTo Configure RoCE on ConnectX-4
- Chelsio-UnifiedWire-Linux-UserGuide on the official website of Chelsio
- Simple NVMe-oF Target Offload Benchmark
- HowTo Configure NVMe over Fabrics (NVMe-oF) Target Offload
- HowTo Configure NVMe over Fabrics Target using nvmetcli
- https://github.com/zrlio/softiwarp